

Enhancing Non-Orthogonal Multiple Access By Forming Relaying Broadcast Channels

Jungho So, *Student Members, IEEE*, and Youngchul Sung[†], *Senior Member, IEEE*

Abstract

In this paper, using relaying broadcast channels (RBCs) as component channels for non-orthogonal multiple access (NOMA) is proposed to enhance the performance of NOMA in single-input single-output (SISO) cellular downlink systems. To analyze the performance of the proposed scheme, an achievable rate region of a RBC with compress-and-forward (CF) relaying is newly derived based on the recent work of noisy network coding (NNC). Based on the analysis of the achievable rate region of a RBC with decode-and-forward (DF) relaying, CF relaying, or CF relaying with dirty-paper coding (DPC) at the transmitter, the overall system performance of NOMA equipped with RBC component channels is investigated. It is shown that NOMA with RBC-DF yields marginal gain and NOMA with RBC-CF/DPC yields drastic gain over the simple NOMA based on broadcast component channels in a practical system setup. By going beyond simple broadcast channel (BC)/successive interference cancellation (SIC) to advanced multi-terminal encoding including DPC and CF/NNC, far larger gains can be obtained for NOMA.

Index Terms

[†]Corresponding author

The authors are with Dept. of Electrical Engineering, KAIST, Daejeon 305-701, South Korea. E-mail: {jhso, yc-sung}@kaist.ac.kr. This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2A10060852).

Relaying broadcast channel, non-orthogonal multiple access, decode-and-forward, compress-and-forward, dirty-paper coding.

I. INTRODUCTION

Motivation: To meet the exponentially growing demand for high data rates in next generation wireless communication systems, enhancing existing lower band wireless systems as well as introducing new bandwidths in higher bands is under vigorous efforts [1]. One of the technologies for increasing the spectral efficiency of cellular systems is recently proposed non-orthogonal multiple access (NOMA) [2], [3]. Traditionally, the wireless communication resources in cellular systems such as time and frequency bandwidth were divided into orthogonal resource blocks, and within a separated resource block only one user is served by the base station (BS). In NOMA, however, multiple users are served non-orthogonally within each resource block by exploiting the power domain. From the system perspective, such user allocation can be regarded as *system overloading* with which the number of served users is larger than that of orthogonal resource blocks. Since multiple users are served non-orthogonally within each resource block with such overloading, the signal from some users allocated to a resource block interferes with other users allocated to the same resource block, but such interference is eliminated by partial user cooperation and non-linear decoding like successive interference cancellation (SIC). For the example of two user allocation in the same resource block, two users with different channel gains are grouped into a resource block so that one user has a higher channel gain (i.e., is close to the BS) and the other user has a lower channel gain (i.e. is far from the BS). Then, the signals of the two users are added and transmitted. At the receiver side, the user close to the BS decodes not only its data but also the data for the user far from the BS, and cancels the signal of the user far from the BS from its received signal. On the other hand, the user far from the BS just decodes its data by treating the interference from the user close to the BS as noise. This is possible due to the asymmetry of the channel gains of the two users since with power

control the user close to the BS requires less power than the user far from the BS and thus the user close to the BS can decode the data intended for the user far from the BS if the user far from the BS can decode its own data. It has been shown that such system overloading based on NOMA can yield non-trivial spectral efficiency increase [2].

From the perspective of information theory, the BS and the unique served user within a separated resource block form a point-to-point (P2P) channel in conventional orthogonalization-based cellular systems. However, under NOMA a broadcast channel (BC) is formed by the BS and the users allocated to the same resource block within a separated resource block. Indeed, it is known that a single-input single-output (SISO) Gaussian BC (GBC) is a degraded BC and the aforementioned super-position coding and SIC achieve its capacity region [4], [5]. Thus, the rate increase by NOMA is due to the change of the channel within each resource block from a P2P channel to a BC since the capacity of a given channel does not change. The penalty for the rate increase is the required cooperation between the served users and the increase in the transmitter and/or receiver side processing.

Some modification has been made to enhance the aforementioned simple NOMA by increasing the level of the cooperation between the served users and changing the type of channel within a resource block [3] with the consideration of the recently available device-to-device (D2D) communication capability [6], [7]. In [3], the authors considered a two-phase (half-duplex) cooperative NOMA in which the BS broadcasts data to both users in the first phase, and the user with good channel helps the other user by transmitting the data for the other user decoded at its site in the first phase to the other user in the second phase. That is, the user with good channel serves as a half-duplex decode-and-forward (DF) relay.* However, such half-duplexing reduces the data rate by half and the resulting system has limitation to increase the system rate.

*In the case of more than two users in a resource block, the same idea can be extended to a multi-phase cooperative NOMA [3].

Summary of Results: In this paper, we further enhance the performance of NOMA by introducing full-duplex relaying at the user with good channel and several relevant encoding schemes at the BS for the case of two-user allocation within each resource block which seems most practical with consideration of performance gain and complexity. When the user with good channel serves as a full-duplex relay, the BS and the two served users form a *relaying broadcast channel (RBC)* from the perspective of information theory [8], [9]. A RBC is different from a BC in that one of the receivers serves as a relay as well as a receiver for its own data, as shown in Fig. 2 in Section III. There exist several known relaying methods such as amplify-and-forward (AF), DF, and compress-and-forward (CF) [10], [11]. In this paper, we consider DF and CF relaying for performance improvement. AF is not relevant in RBCs for NOMA since AF in RBCs amplifies the signal intended to the relaying receiver as well as the signal intended for the other receiver and directly transmits the amplified sum to the other receiver. Several information-theoretical achievable rate region analyses were performed on RBCs. In [8], the authors studied the achievable rate region of a RBC with a DF relaying receiver and showed that the achievable rate region of a RBC subsumes that of the BC generated by eliminating the link between the relaying receiver and the other receiver in the SISO Gaussian case. In [9], the author considered the achievable rate region of a RBC employing CF with common information based on [10]. However, the encoding scheme at the BS proposed in [9] is complicated and does not provide much insight. Furthermore, we are not much interested in the case with common information for both receivers. Thus, we here simplify the problem by eliminating the common information and derive an achievable rate region of a RBC employing CF based on the recent work of noisy network coding (NNC)[†] in [12]. Note that the setup of RBC and that of NNC are different in that a transmitter in NNC has only one message possibly intended for many receivers but in

[†]Noisy network coding for the case of three nodes composed of a transmitter, a relay and a receiver can be viewed as a simplified CF scheme.

RBC the transmitter has two messages intended for two different users. Although the channel setup is different, we still apply the NNC encoding scheme with some modification appropriate to RBC and derive an achievable rate region of a RBC with CF/NNC. Furthermore, based on this result, we derive an achievable rate region of a RBC with CF/NNC when dirty-paper coding (DPC) [13] is applied at the transmitter.

To evaluate the overall system performance of the proposed NOMA with RBC, we consider two user pairing and scheduling methods: near-far pairing and nearest-neighbor pairing. These two pairing methods are opposite to each other and provide two extreme pairing on which the performance of different NOMA schemes can be compared. Based on the achievable rate region result for a RBC with DF in [8] and the newly derived achievable rate region result for a RBC with CF/NNC or CF/NNC/DPC in this paper for each resource block, the overall system performance gain of the proposed NOMA with RBC is examined under the two user pairing and scheduling methods. Numerical results show that the gain of NOMA with RBC-DF is marginal, but NOMA with RBC-CF/NNC/DPC yields drastic gain over the simple NOMA based on GBC/SIC [2] in a practical system setup.

Notations and Organization: We will make use of standard notational conventions. Vectors are written in boldface in lowercase letters. Random variables are written in capitals and the realizations of random variables are written in lowercase letters. For a random variable X , $\mathbb{E}\{X\}$ denotes the expectation of X , and $X \sim \mathcal{CN}(\mu, \Sigma)$ means that X is circularly-symmetric complex Gaussian-distributed with mean μ and covariance Σ .

The remainder of this paper is organized as follows. In Section II, the system model is described. In Section III, the achievable rate region of a RBC is given in general discrete memoryless channel and Gaussian channel cases. The considered user pairing and scheduling methods are described in Section IV. Numerical results are provided in Section V, followed by conclusion in Section VI.

II. SYSTEM MODEL

In this paper, we consider a single-cell SISO downlink system with a single-antenna BS and K single-antenna users, where the considered cell topology is a typical 120° sector of a disk and each user is distributed uniformly in the sector, as shown in Fig. 1. We assume that we have B communication resource blocks that are orthogonal to each other.[‡] The BS selects and assigns M users to each resource block[§] and we assume that $K \geq BM$ to incorporate the impact of multi-user diversity in our system performance investigation. In particular, we focus on the case of $M = 2$ in this paper. Since resource blocks are orthogonal to each other, we can consider

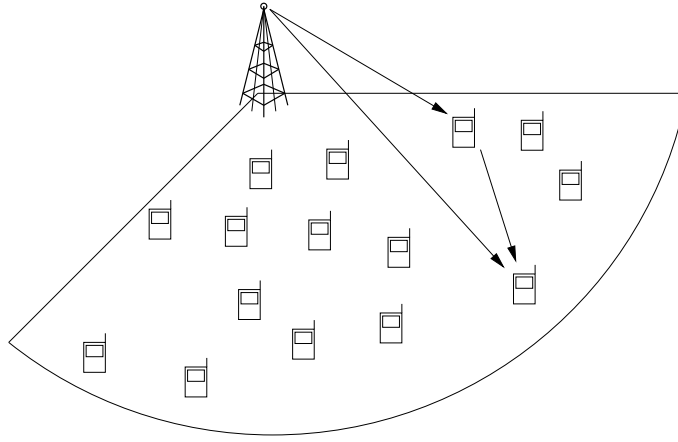


Fig. 1. The considered single-cell SISO downlink system

each resource block separately. Let the indices of the users scheduled to resource block b be 1_b and 2_b with $1_b, 2_b \in \{1, 2, \dots, K\}$, $b = 1, 2, \dots, B$, and let the channel gains from the BS to users 1_b and 2_b be $h_{01_b}^{(b)}$ and $h_{02_b}^{(b)}$, respectively. (Here, $h_{0k}^{(b)}$ is the channel gain from the BS to user k , $k = 1, 2, \dots, K$ at resource block b .) We assume that the indices 1_b and 2_b are ordered

[‡]For example, such resource orthogonalization can be attained by OFDM or other orthogonalization techniques. In the case of OFDM, one resource block represents a subcarrier or a chunk of subcarriers.

[§]The scheduling and grouping method will be explained in Section IV.

such that $|h_{01_b}^{(b)}| \geq |h_{02_b}^{(b)}|$. We assume that the BS and users 1_b and 2_b form a RBC with user 1_b acting as a relaying receiver. Then, the received signals $Y_1^{(b)}$ and $Y_2^{(b)}$ at users 1_b and 2_b in resource block b are respectively given by

$$Y_1^{(b)} = h_{01_b}^{(b)} X_0^{(b)} + Z_1^{(b)}, \quad (1)$$

$$Y_2^{(b)} = h_{02_b}^{(b)} X_0^{(b)} + h_{1_b 2_b}^{(b)} X_1^{(b)} + Z_2^{(b)} \quad (2)$$

where $h_{1_b 2_b}^{(b)}$ represents the channel gain of the link from user 1_b to user 2_b , $X_0^{(b)}$ is the transmit signal at the BS in resource block b with power constraint $\mathbb{E}\{|X_0^{(b)}|^2\} \leq P_0^{(b)}$, $X_1^{(b)}$ is the transmit signal at the relaying user 1_b in resource block b with power constraint $\mathbb{E}\{|X_1^{(b)}|^2\} \leq P_1^{(b)}$, and $Z_1^{(b)} \sim \mathcal{CN}(0, N_1^{(b)})$ and $Z_2^{(b)} \sim \mathcal{CN}(0, N_2^{(b)})$ are the zero-mean additive circularly-symmetric complex Gaussian noise at users 1_b and 2_b in resource block b , respectively. Let the rates of users 1_b and 2_b for resource block b be $R_1^{(b)}$ and $R_2^{(b)}$, respectively. Then, the overall system sum rate R_{sum} is given by

$$R_{sum} = \sum_{i=1}^B (R_1^{(b)} + R_2^{(b)}). \quad (3)$$

The system sum rate R_{sum} is a function of the component channel rates $(R_1^{(b)}, R_2^{(b)})$ for resource block b and the scheduling and grouping method.

III. COMPONENT CHANNEL ANALYSIS: THE RELAYING BROADCAST CHANNEL

In this section, we analyze the achievable rate region of a component RBC composed of the BS and users 1_b and 2_b for each resource block b , which is the backbone for the later stage of this paper. As mentioned already, we consider DF and CF relaying for user 1_b since the relative performance of DF and CF depends on the channel situation but the rate of AF is always worse than the better of DF and CF [11]. Note that a RBC is different from a relay channel since the transmitter sends two information messages: one for the relaying receiver and the other for the other receiver. We shall call RBC with DF and CF RBC-DF and RBC-CF, respectively. In the following subsections, we investigate the achievable rate regions of RBC-DF and RBC-CF

in the discrete memoryless channel case first. Based on the result in the discrete memoryless channel case, we obtain the achievable rate regions of RBC-DF and RBC-CF in the Gaussian channel case next.

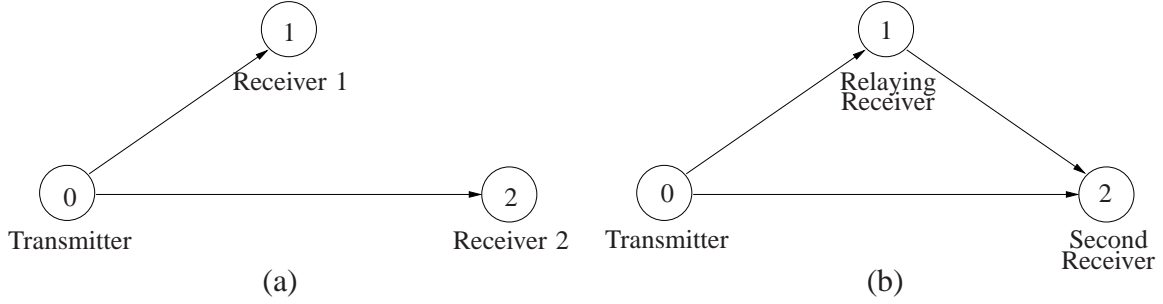


Fig. 2. (a) a broadcast channel (BC) and (b) a relaying broadcast channel (RBC)

A. The Discrete Memoryless Case

A general RBC is a 3-node discrete memoryless network composed of node 0 (the transmitter), node 1 (called the relaying receiver), and node 2 (called the second receiver), as depicted in Fig. 2(b), defined by

$$(\mathcal{X}_0 \times \mathcal{X}_1, p(y_1, y_2|x_0, x_1), \mathcal{Y}_1 \times \mathcal{Y}_2), \quad (4)$$

where \mathcal{X}_0 and \mathcal{X}_1 are the input alphabets of nodes 0 and 1, respectively; \mathcal{Y}_1 and \mathcal{Y}_2 are the output alphabets of nodes 1 and 2, respectively; and $p(y_1, y_2|x_0, x_1)$ is the channel transition probability mass function.

From here on, we investigate the achievable rate region of the considered RBC. First, we consider the case that node 1 does not transmit signal to node 2, i.e. $\mathcal{X}_1 = \emptyset$. Then, the channel reduces to a 2-user BC and the capacity region of a degraded BC is given by the following theorem.

Theorem 1: [5] The capacity region of the degraded discrete memoryless BC $(\mathcal{X}_0, p(y_1, y_2|x_0), \mathcal{Y}_1 \times \mathcal{Y}_2)$ is the set of rate pairs (R_1, R_2) such that

$$R_1 < I(X_0; Y_1|U) \quad (5)$$

$$R_2 < I(U; Y_2) \quad (6)$$

for some pmf $p(u, x_0)$, where the cardinality of the auxiliary random variable U satisfies $|\mathcal{U}| \leq \min\{|\mathcal{X}_0|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\} + 1$. Here, R_1 and R_2 are the rates of nodes 1 and 2, respectively.

It is known that the capacity region of a degraded BC can be achieved by superposition coding and SIC. This can be seen in the rate formulae (5) and (6). Here, the auxiliary random variable U is associated with the message to node 2. In (5), R_1 is bounded by the mutual information between the transmitted signal variable X_0 and node 1's received signal variable Y_1 conditioned on U . Conditioning can be viewed interference cancellation and means that node 1 decodes the message associated with node 2. On the other hand, R_2 is bounded simply by the mutual information between its message variable U and its received signal variable Y_2 . The above capacity region result is used in the simple NOMA [2].

Now, consider the RBC scheme with DF at node 1. In this case, contrary to the 2-user BC, we have $\mathcal{X}_1 \neq \emptyset$, which means that node 1 not only decodes the data for itself but also actively helps node 2. When the DF relaying scheme is applied to the considered RBC, we have the following rate region result given by [8]:

Theorem 2: [8] The rate pair (R_1, R_2) is achievable for the RBC (4) if

$$R_1 < I(X_0; Y_1|U, X_1) \quad (7)$$

$$R_2 < \min\{I(U; Y_1|X_1), I(U, X_1; Y_2)\} \quad (8)$$

for some joint distribution $p(x_1)p(u|x_1)p(x_0|u)$, where U is an auxiliary random variable associated with the message for node 2.

The achievable rate region in Theorem 2 can also be obtained by using superposition coding at node 0 and SIC at node 1 similarly to the result in Theorem 1 and in addition by node 1's transmitting the decoded (at node 1) data (intended for node 2) to node 2. In Theorem 2, U is an auxiliary random variable associated with the message for node 2 and X_0 is the input random variable at node 0 associated with the messages for both nodes 1 and 2. In (7), conditioning on U means that node 1 decodes the message associated with node 2, and then the mutual information between X_0 and Y_1 conditioned on U is related to the rate of the message for node 1. (Here, node 1 knows its own transmit variable X_1 and thus node 1 can cancel the self-interference. This is seen as conditioning on X_1 in (7).) Regarding (8), the first term in the right-hand side (RHS) in (8) means that the message intended for node 2 should be decoded successfully at node 1 for DF operation and the second term in the RHS in (8) is related to the rate at which node 2 decodes its message (U) based on its received signal Y_2 with the help (X_1) from node 1. Taking minimum in (8) means both events should happen in this scheme. Note that if we remove X_1 , (7) reduces to (5), and the second term in the RHS of (8) reduces to (6). Here, the first term $I(U; Y_1)$ in the RHS of (8) without X_1 is always larger than or equal to the second term $I(U; Y_2)$ in the RHS of (8) without X_1 , i.e., $I(U; Y_1) \geq I(U; Y_2)$ due to the assumption of degradedness $U \rightarrow Y_1 \rightarrow Y_2$. Therefore, the achievable rate region in Theorem 2 always subsumes the capacity region in Theorem 1. In other words, NOMA with the proposed RBC-DF is always better than the simple NOMA adopting the degraded BC as its component channel in [2].

When node 1 can decode the data intended for node 2, RBC-DF always performs better than RBC-AF since the correct message for node 2 is regenerated at node 1 and forwarded to node 2, but in RBC-AF node 1 only forwards a noise-corrupted version of the message for node 2 directly to node 2. Note the rate R_2 in (8) for RBC-DF is limited by the term $I(U; Y_1|X_1)$ resulting from the requirement that node 2's message should be decoded successfully at node 1 for DF operation. One way to circumvent this full decoding requirement is the CF scheme in

which node 2's information is compressed at node 1 and forwarded to node 2 [10]. It is known that CF outperforms DF under certain situations [11]. When full decoding of node 2's data at node 1 is not possible or results in a low rate, we can resort to RBC-CF. Furthermore, RBC-CF performs better than RBC-AF since AF is worse than CF [11]. Thus, we consider RBC-CF adopting CF at node 1 as our next choice for the component channel. An achievable rate region of a RBC-CF with common information intended for both nodes 1 and 2 was derived in [9]. However, the derivation and the encoding scheme are complicated and do not provide much insight. Hence, we here simplify the problem by eliminating common information and derive a simple achievable rate region of a RBC-CF based on the recent encoding and compression technique of noisy network coding (NNC) presented in [12]. The NNC in the 3-node setup is a simplified CF scheme compared to the original CF scheme proposed in [10]. Although we use the coding technique in NNC, there is a fundamental difference between NNC and RBC. In the NNC setup, the transmitter has only one message which may be intended for multiple receivers. In RBC, however, the transmitter has two messages: one for the relaying receiver and the other for the second receiver. By extending the NNC scheme to RBC, we obtain the following result regarding the achievable rate region of a RBC.

Theorem 3: The rate pair (R_1, R_2) is achievable for the RBC (4) if

$$R_1 < I(U; Y_1) \quad (9)$$

$$R_2 < \min\{I(V; \hat{Y}_1, Y_2 | X_1), I(V, X_1; Y_2) - I(\hat{Y}_1; Y_1 | V, X_1, Y_2)\} \quad (10)$$

for some joint distribution

$$p(x_1)p(u)p(v)p(x_0|u, v)p(\hat{y}_1|y_1, x_1). \quad (11)$$

Proof: See Appendix A. ■

Here, U and V are the input message variables to node 1 (the relaying receiver) and node 2 (the second receiver), respectively, and the overall transmit variable X_0 of node 0 is generated based

on (U, V) , as seen in the term $p(x_0|u, v)$ in the generating input distribution in (11). Thus, the rate R_1 is simply the mutual information between the message variable U for node 1 and the received signal variable Y_1 at node 1. The rate R_2 is the rate of NNC with V as the transmit variable at node 0, where the cut-set bound is used [5], [12]. The first term $I(V; \hat{Y}_1, Y_2|X_1)$ in the RHS of (10) is the mutual information between node 0 and nodes $\{1, 2\}$ with self interference cancellation at the cut group, nodes $\{1, 2\}$. The term $I(V, X_1; Y_2)$ [¶] in the second term in the RHS of (10) is the decoding rate of node 2 with the help (X_1) from node 1 and the term $I(\hat{Y}_1; Y_1|V, X_1, Y_2)$ in the second term in the RHS of (10) represents the loss related to compression compared to full decoding. For the details of the encoding and decoding scheme for the rate-tuple in Theorem 3, see Appendix A.

B. The Gaussian Case

In this section, we consider the Gaussian channel case and compare the performance of the three component channel formulation schemes: GBC (simple NOMA), RBC-DF, and RBC-CF/NNC. In the Gaussian channel case, the received signals at the relaying receiver and the second receiver are given by (1) and (2), respectively, which are rewritten here as

$$Y_1 = h_{01}X_0 + Z_1, \quad (12)$$

$$Y_2 = h_{02}X_0 + h_{12}X_1 + Z_2, \quad (13)$$

where the resource block superscript (b) is omitted. Here, Y_1 and Y_2 are the received signals at the relaying receiver and the second receiver, respectively; X_0 and X_1 are the transmit signals from the transmitter and the relaying receiver, respectively; h_{ij} denotes the channel from node i to node j ; and $Z_i \sim \mathcal{CN}(0, N_i)$ is the zero-mean additive Gaussian noise at node i .

To compute the rate-tuples in Theorems 1, 2, and 3, we need to specify the associated input distributions since the channel $p(y_1, y_2|x_0, x_1)$ is given. We set $X_0 \sim \mathcal{CN}(0, P_0)$ and $X_1 \sim$

[¶]This term corresponds to the second term $I(U, X_1; Y_2)$ in the RHS of (8) in the RBC-DF scheme.

$\mathcal{CN}(0, P_1)$, and set the transmitted signal at node 0 (i.e., the transmitter) as the superimposed signal given by

$$X_0 = U + V, \quad (14)$$

where U is the signal for node 1 and V is the signal for node 2:

$$U \sim \mathcal{CN}(0, \alpha P_0), \quad V \sim \mathcal{CN}(0, \bar{\alpha} P_0), \quad \bar{\alpha} = 1 - \alpha, \quad 0 \leq \alpha \leq 1. \quad (15)$$

It is known that a two-user SISO GBC is a degraded BC since either $\frac{|h_{01}|^2}{N_1} \geq \frac{|h_{02}|^2}{N_2}$ or $\frac{|h_{01}|^2}{N_1} < \frac{|h_{02}|^2}{N_2}$. With the considered ordering in Section II, we have $\frac{|h_{01}|^2}{N_1} \geq \frac{|h_{02}|^2}{N_2}$. Then, the following NOMA condition is automatically satisfied:

$$\frac{|h_{01}|^2 \bar{\alpha} P_0}{|h_{01}|^2 \alpha P_0 + N_1} \geq \frac{|h_{02}|^2 \bar{\alpha} P_0}{|h_{02}|^2 \alpha P_0 + N_2}. \quad (16)$$

The capacity region of GBC (simple NOMA) is given by

$$R_1 \leq \log \left(1 + \frac{|h_{01}|^2 \alpha P_0}{N_1} \right), \quad (17)$$

$$R_2 \leq \log \left(1 + \frac{|h_{02}|^2 \bar{\alpha} P_0}{|h_{02}|^2 \alpha P_0 + N_2} \right). \quad (18)$$

Next, consider the RBC-DF scheme. From Theorem 2, the achievable rate region is given by

$$R_1 \leq \log \left(1 + \frac{|h_{01}|^2 \alpha P_0}{N_1} \right), \quad (19)$$

$$R_2 \leq \min \left\{ \log \left(1 + \frac{|h_{02}|^2 \bar{\alpha} P_0 + |h_{12}|^2 P_1}{|h_{02}|^2 \alpha P_0 + N_2} \right), \log \left(1 + \frac{|h_{01}|^2 \bar{\alpha} P_0}{|h_{01}|^2 \alpha P_0 + N_1} \right) \right\}. \quad (20)$$

From the fact that the rates (17) and (19) for R_1 are the same and (18) for R_2 is always smaller than or equal to (20) for R_2 by the condition (16), we can easily see that the achievable rate region of RBC-DF subsumes the capacity region of simple NOMA based on GBC. The improvement of rate R_2 is large when $|h_{12}|^2 P_1$ is large and the gap between $\frac{|h_{01}|^2}{N_1}$ and $\frac{|h_{02}|^2}{N_2}$ is large.

Now, consider RBF-CF/NNC in the Gaussian case. Here we use Theorem 3 to derive an achievable rate region in the Gaussian case. Note that the input distribution in this case is given

by $p(x_1)p(u)p(v)p(x_0|u, v)p(\hat{y}_1|y_1, x_1)$ in (11). Thus, to apply Theorem 3 to the Gaussian channel case, we further set the remaining part $p(\hat{y}_1|y_1, x_1)$ of the input distribution as

$$\hat{Y}_1 = Y_1 - h_{01}U + \hat{Z} = h_{01}V + Z_1 + \hat{Z}, \quad (21)$$

where $\hat{Z} \sim \mathcal{CN}(0, \hat{N})$. With some calculation, we get the following achievable rate region of RBC-CF/NNC:

$$R_1 \leq \log \left(1 + \frac{|h_{01}|^2 \alpha P_0}{|h_{01}|^2 \bar{\alpha} P_0 + N_1} \right), \quad (22)$$

$$R_2 \leq \min \left\{ \log \left(1 + \frac{(|h_{02}|^2 \alpha P_0 + N_2)|h_{01}|^2 \bar{\alpha} P_0 + (N_1 + \hat{N})|h_{02}|^2 \bar{\alpha} P_0}{(N_1 + \hat{N})(|h_{02}|^2 \alpha P_0 + N_2)} \right), \right. \\ \log \left(1 + \frac{|h_{02}|^2 \bar{\alpha} P_0 + |h_{12}|^2 P_1}{|h_{02}|^2 \alpha P_0 + N_2} \right) \\ \left. - \log \left(1 + \frac{N_1^2 N_2 + N_1^2 |h_{02}|^2 \alpha P_0}{\hat{N} N_1 N_2 + \hat{N} N_2 |h_{01}|^2 \alpha P_0 + \hat{N} N_1 |h_{02}|^2 \alpha P_0 + N_1 N_2 |h_{01}|^2 \alpha P_0} \right) \right\} \quad (23)$$

(The detail of the calculation is in Appendix B.) The rate R_2 of the second receiver in (23) can be larger than that of RBC-DF in (20) depending on the situation. However, note that the rate R_1 of the relaying receiver in (22) is smaller than that of GBC and RBC-DF. This is because at the relaying receiver the message for the second receiver is not fully decoded and thus the interference from the second receiver's signal at the relaying receiver cannot be cancelled by SIC. To resolve this problem, we apply DPC [13] at the transmitter together with the encoding scheme presented in Theorem 3 to remove the interference from the second receiver's signal at the relaying receiver since the transmitter knows both messages [14]. In this case, the transmitter generates the message codeword for the second receiver first and then based on this message codeword it generates the message codeword for the relaying receiver based on DPC. Then, the transmitter superimposes the two message codewords and transmits the superimposed signal. The processing at the relaying receiver and the second receiver is the same as RBC-CF/NNC. In the decoding process of the relaying receiver for its own message, the interference from the second receiver's signal is automatically removed due to DPC applied at the transmitter side.

The achievable rate region of RBC-CF/NNC employing DPC is given by

$$R_1 \leq \log \left(1 + \frac{|h_{01}|^2 \alpha P_0}{N_1} \right) \quad (24)$$

$$R_2 \leq \min \left\{ \log \left(1 + \frac{(|h_{02}|^2 \alpha P_0 + N_2)|h_{01}|^2 \bar{\alpha} P_0 + (N_1 + \hat{N})|h_{02}|^2 \bar{\alpha} P_0}{(N_1 + \hat{N})(|h_{02}|^2 \alpha P_0 + N_2)} \right), \right. \\ \log \left(1 + \frac{|h_{02}|^2 \bar{\alpha} P_0 + |h_{12}|^2 P_1}{|h_{02}|^2 \alpha P_0 + N_2} \right) \\ \left. - \log \left(1 + \frac{N_1^2 N_2 + N_1^2 |h_{02}|^2 \alpha P_0}{\hat{N} N_1 N_2 + \hat{N} N_2 |h_{01}|^2 \alpha P_0 + \hat{N} N_1 |h_{02}|^2 \alpha P_0 + N_1 N_2 |h_{01}|^2 \alpha P_0} \right) \right\}. \quad (25)$$

Note that in this scheme R_2 is the same as (23) of RBC-CF/NNC but R_1 is improved to be the same as (19) of GBC and RBC-DF. The value of \hat{N} can be optimized to yield maximum R_2 in (23) and (25) by solving a quadratic equation. The proposed encoding scheme based on both superposition/DPC and CF/NNC for NOMA is described in Fig. 5 in Appendix A.

IV. THE CONSIDERED USER SCHEDULING AND PAIRING

In Section III, we have investigated the achievable regions for several component channel types. In this section, we introduce two user pairing methods to compare the performance of the overall system adopting one of the considered component channel types: GBC (simple NOMA), RBC-DF or RBC-CF as the component channel. Since the performance of the overall system depends on user pairing, we consider two disparate user pairing methods: near-far pairing and nearest neighbor pairing. The two pairing methods are opposite to each other and are useful to compare NOMA employing a different component channel type in different system setting.

A. Near-Far Pairing

The first considered user scheduling and pairing is similar to that in [15] except that we consider a sequential approach. In the first method, we aim at pairing two users: one with good channel and the other with bad channel. We assume that the power for the relaying receiver and the power for the second receiver for each resource block are fixed, i.e., the parameter α in (15)

is given, and the BS knows the location of each user in the cell and the gain of the channel from the BS itself to each user in the cell, i.e., $h_{0k}^{(b)}$ for $k = 1, 2, \dots, K$ and $b = 1, 2, \dots, B$. First, the users in the cell are divided into two groups for each resource block b : group $G_1^{(b)}$ with good channel with $K/2$ users and group $G_2^{(b)}$ with bad channel with $K/2$ users by ordering $|h_{0k}^{(b)}|$ for each resource block b . Then, for resource block $b = 1$, we pick one user (which becomes the relaying receiver) from $G_1^{(1)}$ based on the proportionally fair (PF) scheduling [16] and the instantaneous achievable rate R_1 given in Section III-B for RBC-DF, RBC-CF/NNC, or RBC-CF/NNC/DPC. That is, the selected user is given by

$$\kappa_1^{(1)} = \arg \max_{i \in G_1^{(1)}} \frac{R_{1(i)}^{(b)}[t]}{\bar{\mathcal{R}}(i)[t]}, \quad (26)$$

where $R_{1(i)}^{(b)}[t]$ is the rate R_1 given in Section III-B when user i serves as the relaying receiver at time t and resource block b , and $\bar{\mathcal{R}}(i)[t]$ is the average served rate for user i up to time t . Note from Section III-B that $R_{1(i)}^{(b)}[t]$ can be computed based only on $h_{0i}^{(b)}$. After $\kappa_1^{(1)}$ is chosen, we select the second user $\kappa_2^{(1)}$ for resource block $b = 1$ from $G_2^{(1)}$ based on $\kappa_1^{(1)}$ and again the PF principle, i.e.,

$$\kappa_2^{(1)} = \arg \max_{i \in G_2^{(1)}} \frac{R_{2(i|\kappa_1^{(1)})}^{(b)}[t]}{\bar{\mathcal{R}}(i)[t]}, \quad (27)$$

where $R_{2(i|\kappa_1^{(1)})}^{(b)}[t]$ is the rate R_2 given in Section III-B when user i is the second receiver paired with the relaying receiver j at time t and resource block b . Here, as seen in Section III-B, the computation of $R_{2(i|\kappa_1^{(1)})}^{(b)}[t]$ requires the knowledge of the channel gain $|h_{\kappa_1^{(1)}i}^{(b)}|^2$ from user $\kappa_1^{(1)}$ and user i . In this step, we use an estimate for the channel gain based on [17]

$$\widehat{|h_{ij}^{(b)}|^2} = C_0 d^{-\gamma}, \quad (28)$$

where C_0 is a constant, d is the distance between users i and j , and γ is the path loss exponent. (The assumption of knowledge of user locations at the BS is required for this step.) After $\kappa_1^{(1)}$ and $\kappa_2^{(1)}$ for resource block $b = 1$ are selected, we proceed to $b = 2$. For resource block $b = 2$,

we remove $\kappa_1^{(1)}$ and $\kappa_2^{(1)}$ from $G_1^{(2)}$ and $G_2^{(2)}$, and repeat the same procedure with the remaining sets. After users are selected for all resource blocks, we update the average served rate for the served users as

$$\bar{\mathcal{R}}(i)[t+1] := (1-\tau)\bar{\mathcal{R}}(i)[t] + \tau R(i)[t], \quad i = 1, \dots, K, \quad (29)$$

where $R(i)[t]$ is the served rate for user i at time t , and τ is the auto-regressive (AR) filter coefficient or forgetting factor.

B. Nearest-Neighbor Pairing

The second scheduling and pairing is quite opposite to the first method. In the second method, we aim at pairing two users who are close to each other. The reason of considering the second pairing method is to investigate the performance of general NOMA over a wide range of user pairing methods. In the second method, we select one user as the relaying receiver and its nearest neighbor as the second receiver. Since the nearest neighbor for each user is given, we can select the two users simultaneously based on the PF metric. That is, for resource block $b = 1$, set $\mathcal{G} = \{1, 2, \dots, K\}$ and

$$\kappa_1^{(1)} = \arg \max_{i \in \mathcal{G}} \left(\frac{R_{1(i)}^{(b)}[t]}{\bar{\mathcal{R}}(i)[t]} + \frac{R_{2(\mathcal{N}(i)|i)}^{(b)}[t]}{\bar{\mathcal{R}}(\mathcal{N}(i))[t]} \right) \quad (30)$$

where $\mathcal{N}(i)$ is the index of the nearest neighbor of user i . When user selection for resource block $b = 1$ is finished, we remove $\kappa_1^{(1)}$ and $\kappa_2^{(1)} = \mathcal{N}(\kappa_1^{(1)})$ from set \mathcal{G} , and repeat the same procedure for resource blocks $b = 2, \dots, B$.

V. NUMERICAL RESULTS

In this section, we provide some numerical results to evaluate the performance of the proposed NOMA with RBC. We first evaluate the performance of each component channel presented in Section III-B and then evaluate the sum rate of the entire cell employing the considered user pairing and scheduling presented in Section IV and the considered RBC component channel.

A. The Component Channel Performance

For the evaluation of the performance of component channels, we considered a linear configuration in which the location of the relaying receiver in the middle of the line between the transmitter and the second receiver. We set $N_1 = N_2 = N = 1$ and considered three pairs of $(P_0^{(b)}, P_1^{(b)})$ for the transmit power $P_0^{(b)}$ at the BS and the transmit power $P_1^{(b)}$ at the relaying receiver: $(P_0^{(b)}/N, P_1^{(b)}/N) = (10 \text{ dB}, 10 \text{ dB}), (10 \text{ dB}, 5 \text{ dB}), \text{ and } (10 \text{ dB}, 0 \text{ dB})$.^{||} We assumed that the path loss exponent is $\gamma = 3$. Based on $\gamma = 3$, we considered two channel gain setup: (i)** $|h_{01}|^2 = |h_{12}|^2 = 8$ and $|h_{02}|^2 = 1$ and (ii) $|h_{01}|^2 = |h_{12}|^2 = 1$ and $|h_{02}|^2 = 1/8$. Then, we swept the value of the parameter α defined in (15) to determine the achievable rate pair (R_1, R_2) . The result is shown in Fig. 3. Fig.3(a), (c) and (e) show the rate-tuples in [(17), (18): GBC - simple NOMA], [(19), (20): RBC-DF], [(22), (23): RBC-CF/NNC], and [(24), (25): RBC-CF/NNC/DPC] for the channel gain setting (i) of around 10 dB received SNR operation. It is seen that the proposed NOMA equipped with RBC component channels employing superposition/DPC and CF/NNC significantly improves the performance over the simple NOMA based on GBC/SIC. The marked points in Fig. 3 are the rate-pair points of $\alpha = 0.2$. It is seen that for $\alpha = 0.2$, R_2 of RBC-CF/NNC without DPC is higher than R_2 of RBC-DF but R_1 of RBC-CF/NNC without DPC is much lower than R_2 of RBC-DF, as expected. It is also seen that in the channel gain setting (i) of roughly 10 dB received SNR operation, the gain of RBC-DF over GBC is not so

^{||}In real-world cellular systems, the maximum BS downlink average transmit power is 43 dBm (20W) and the maximum average transmit power of a cellular phone is 24 dBm (0.25W). However, the BS downlink transmit power is shared by 50 to 100 simultaneous users. Hence, the maximum per-user BS downlink average power is around 23 dBm to 26 dBm. This is the basis for the consider relative magnitude for $P_0^{(b)}$ and $P_1^{(b)}$.

^{**}With the channel gain setting (i), we have node 1's received signal-to-noise ratio (SNR) $|h_{01}|^2 \alpha P_0^{(b)}/N_1 = 12 \text{ dB}$ and node 2's received SNR $|h_{02}|^2 (1 - \alpha) P_0^{(b)}/N_2 = 9 \text{ dB}$ for $\alpha = 0.2$, a typical power distribution value in NOMA [15]. Node 2's SNR of 9dB is higher than the signal-to-interference ratio (SIR) of $0.8/0.2=6 \text{ dB}$. With the channel gain setting (ii), each node's SNR is reduced by 9dB.

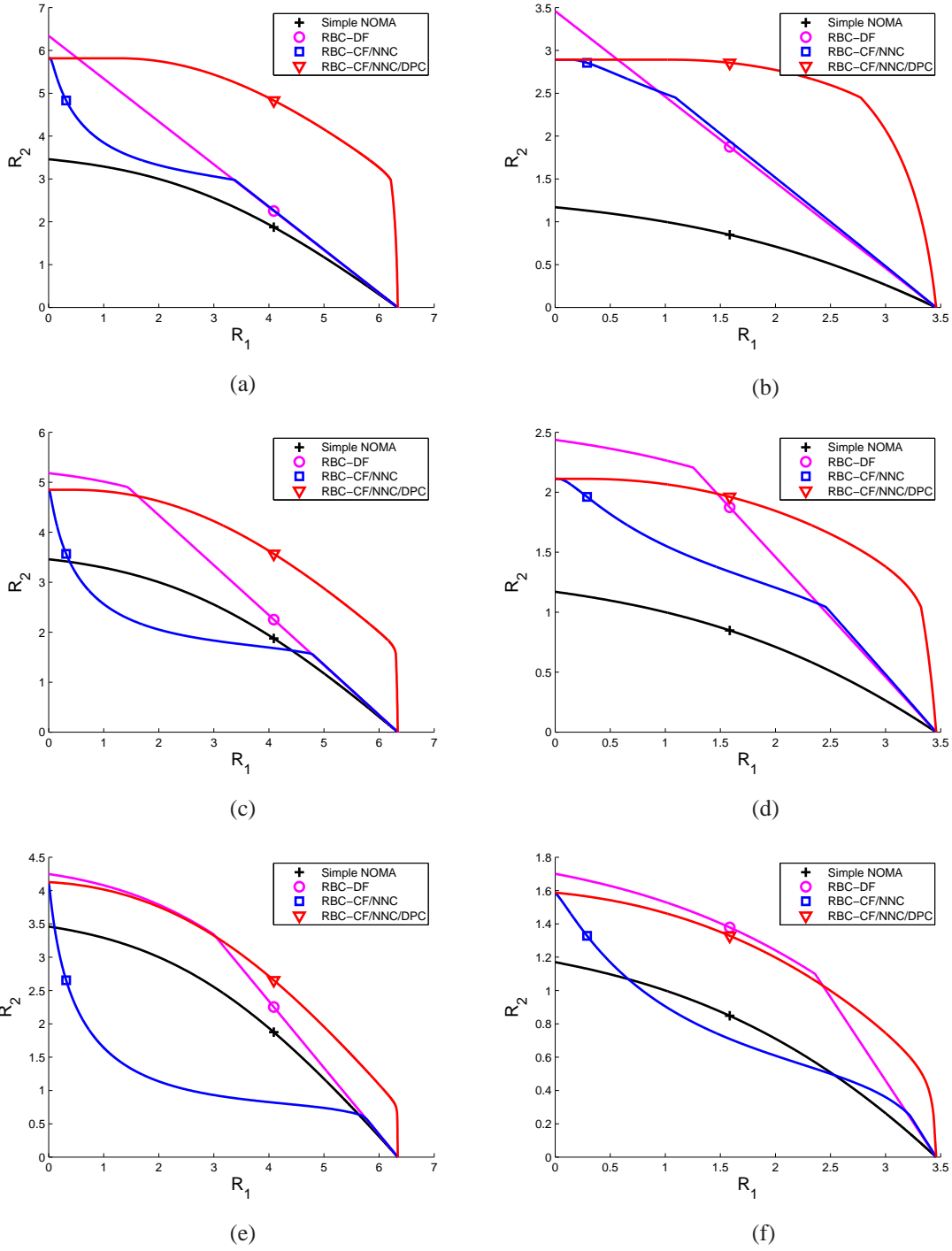


Fig. 3. The achievable rate region - $(|h_{01}|^2, |h_{02}|^2, |h_{12}|^2) = (8, 8, 1)$: (a), (c) and (e), $(|h_{01}|^2, |h_{02}|^2, |h_{12}|^2) = (1, 1, 1/8)$: (b), (d) and (f). $N = N_1 = N_2 = 1$. $(P_0^{(b)}/N, P_1^{(b)}/N) = (10\text{dB}, 10\text{dB})$: (a) and (b), $(P_0^{(b)}/N, P_1^{(b)}/N) = (10\text{dB}, 5\text{dB})$: (c) and (d), $(P_0^{(b)}/N, P_1^{(b)}/N) = (10\text{dB}, 0\text{dB})$: (e) and (f)

large at $\alpha = 0.2$. Fig.3(b), (d) and (f) show the rate-tuples in the channel gain setting (ii) of 0dB received SNR operation. It is seen that the gain by the RBC-CF/NNC/DPC over the simple NOMA (GBC) is drastic.

B. The Overall System Performance

Here, we provide numerical results to evaluate the overall system performance of NOMA with each of the proposed component channels based on the considered user scheduling and pairing method in Section IV in a single-cell downlink network with the cell topology described in Fig. 1. The sector radius from the BS to the cell edge was set to be $D_e = 500$ m. We considered $B = 4$ resource blocks and $K = 40$ users uniformly distributed over the 120° sector from radius 50 m to the cell edge. The noise power for each user was the same and set to be $N = N_1 = N_2 = \dots = N_K = 1$. The channel gain $h_{0k}^{(b)}$ from the BS to user k at the resource block b was modelled as the product of a Rayleigh fading factor $f_{0k}^{(b)} \stackrel{i.i.d.}{\sim} \mathcal{CN}(0, 1)$ and the path loss, given by

$$h_{0k}^{(b)} = f_{0k}^{(b)} \cdot \left(\frac{d_{0k}}{D_e} \right)^{-\gamma}, \quad (31)$$

where d_{0k} was the distance from the BS to user k and the path loss factor was $\gamma = 3$. The BS transmit power $P_0^{(b)}$ was set so that the expected received SNR at the cell edge was 10 dB, i.e.,

$$10\text{dB} = \frac{\mathbb{E}\{|h_{0k}^{(b)}|^2\} P_0^{(b)}}{N} = \frac{\mathbb{E}\{|f_{0k}^{(b)}|^2\} \left(\frac{D_e}{d_{0k}} \right)^{-3} P_0^{(b)}}{N} = \frac{P_0^{(b)}}{N} \quad \forall b = 1, \dots, B.$$

Thus, users with $d_{0k} < D_e$ had expected SNR larger than 10 dB. The transmit power $P_1^{(b)}$ of the relaying receiver was set relative to $P_0^{(b)}$. For one realization of user locations, we ran the user scheduling and pairing method in Section IV with the PF forgetting factor $\tau = 0.01$ in (29) for 1000 scheduling intervals, and computed the sum rate divided by 1000 for each scheme. We averaged the sum rate over 50 independent realizations for user locations. Fig. 4 shows the sum rate result for NOMA equipped with four different component channels: GBC

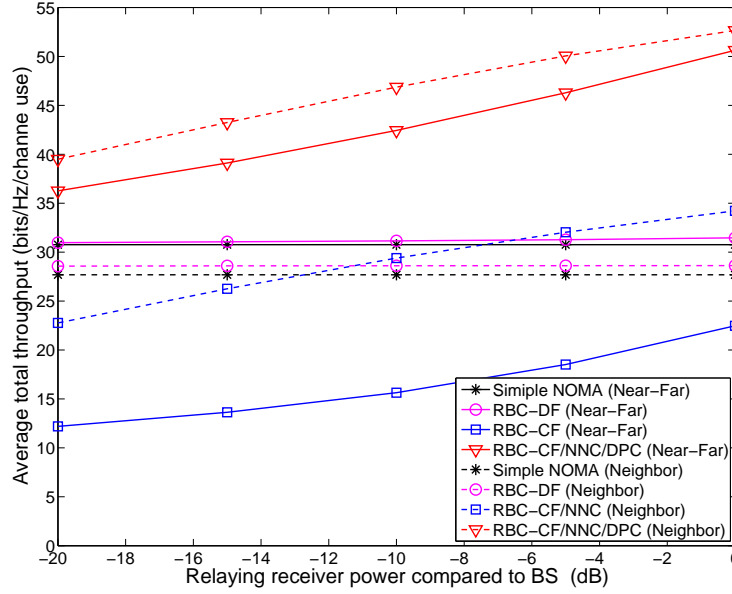


Fig. 4. Total system sum rate: solid line - the near-far pairing and dashed line - the nearest neighbor pairing

(simple NOMA), RBC-DF, RBC-CF/NNC, and RBC-CF/NNC/DPC. For the solid lines the near-far pairing was used and for the dashed lines the nearest neighbor pairing was used. It is seen that the gain by RBC-DF is marginal in this operating SNR range with the cell-edge user SNR of 10 dB, as expected from Section V-A. It is seen that the gain of RBC-CF/NNC/DPC over the simple NOMA is significant when $P_1^{(b)}$ is comparable to $P_0^{(b)}$, as expected from Section V-A. If the operating SNR is decreased, then the gain of NOMA based on RBC will increase further, as expected from Fig. 3(b), (d), and (f). Note that the performance difference due to the two disparate user pairing methods is not so significant for GBC (simple NOMA) and RBC-DF.

VI. CONCLUSION

In this paper, we have considered enhancing NOMA by using RBC component channels in SISO cellular downlink systems. We have newly derived an achievable rate region of a RBC with CF/NNC and have investigated the achievable rate region of a RBC with DF, CF/NNC,

and CF/NNC plus DPC. Based on the achievable rate analysis, we have investigated the overall system performance of NOMA equipped with RBC component channels, and have shown that NOMA with RBC-DF yields marginal gain and NOMA with RBC-CF/NNC/DPC yields drastic gain over the simple NOMA based on GBC in a practical system setup. The gist of the gain of NOMA lies in non-linear processing to cope with system overloading. By going beyond simple GBC/SIC to advanced multi-terminal encoding including DPC and CF/NNC, we can obtain far larger gains. Currently, active research is going on to implement practical DPC and CF codes already with some available codes [18]–[25]. With reflecting the gain in NOMA by using such multi-terminal encoding, it is worth considering such advanced multi-terminal encoding for NOMA.

APPENDIX A

PROOF OF THEOREM 3

Codebook Generation: Fix $p(x_1)p(u)p(v)p(x_0|u, v)p(\hat{y}_1|y_1, x_1)$. We assume blockwise^{††} transmission with n code symbols as one block, and transmit J blocks. We randomly and independently generate a codebook for each block. For each block $j \in [1 : J] \triangleq \{1, 2, \dots, J\}$,

- randomly and independently generate $2^{n\hat{R}_2}$ sequences $\mathbf{x}_{1j}(l_{j-1})$, $l_{j-1} \in [1 : 2^{n\hat{R}_2}]$, each according to the distribution $\prod_{k=1}^n p_{X_1}(x_{1,(j-1)n+k})$;
- randomly and independently generate 2^{nR_1} sequences $\mathbf{u}_j(m_{1j})$, $m_{1j} \in [1 : 2^{nR_1}]$, each according to $\prod_{i=1}^n p_U(u_{(j-1)n+i})$;
- randomly and independently generate 2^{nJR_2} sequences $\mathbf{v}_{1j}(m_2)$, $m_2 \in [1 : 2^{nJR_2}]$, each according to the distribution $\prod_{k=1}^n p_V(v_{(j-1)n+k})$;
- for each $\mathbf{u}_j(m_{1j})$ and $\mathbf{v}_j(m_2)$, randomly generate a sequence $\mathbf{x}_{0j}(m_{1j}, m_2)$ each according to $\prod_{i=1}^n p_{X|U,V}(x_{0,(j-1)n+i}|u_{(j-1)n+i}(m_{1j}), v_{(j-1)n+i}(m_2))$; and

^{††}The term 'block' in the appendix is not the resource block in the main content of the paper. A block in this appendix is a concatenation of n channel code symbols.

- for each $\mathbf{x}_{1j}(l_{j-1})$, randomly and conditionally independently generate $2^{n\hat{R}_2}$ sequences $\hat{\mathbf{y}}_{1j}(l_j|l_{j-1})$, $l_j \in [1 : 2^{n\hat{R}_2}]$, each according to $\prod_{i=1}^n p_{\hat{Y}_1|X_1}(\hat{y}_{1,(j-1)n+i}|x_{1,(j-1)n+i}(l_{j-1}))$.

Then, the codebook is shared for all nodes. The Markov chain relationship between the codewords $(\mathbf{x}_{1j}, \mathbf{u}_j, \mathbf{v}_j, \mathbf{x}_{0j}, \text{ and } \hat{\mathbf{y}}_{1j})$ and the received signal vectors $(\mathbf{y}_{1j} \text{ and } \mathbf{y}_{2j})$ is described in Fig. 5.

Encoding: Let m_{1j} and m_2 be the messages to be sent, and choose $l_0 = 1$ by convention. The transmitter sends $\mathbf{x}_{0j}(m_{1j}, m_2)$ generated from $\mathbf{u}_j(m_{1j})$ and $\mathbf{v}_j(m_2)$.

Upon reception of \mathbf{y}_{1j} , the relaying receiver finds an index l_j such that

$$(\hat{\mathbf{y}}_{1j}(l_j|l_{j-1}), \mathbf{y}_{1j}, \mathbf{x}_{1j}(l_{j-1})) \in \mathcal{T}_{\epsilon_1}^{(n)}(\hat{Y}_1, Y_1, X_1), \quad (32)$$

where $\mathcal{T}_{\epsilon_1}^{(n)}(\hat{Y}_1, Y_1, X_1)$ is the set of ϵ_1 -jointly typical sequences. If there are more than one such index, choose one of them arbitrarily. If there is no such index, choose an arbitrary index. By the covering lemma [5], if $\hat{R}_2 > I(\hat{Y}_1; Y_1|X_1) + \delta_1(\epsilon_1)$, the probability that there exists at least one such index tends to 1 as $n \rightarrow \infty$, where $\epsilon_1 > 0$ and $\delta_1(\cdot)$ is a positive function such that $\delta_1(\epsilon_1) \rightarrow 0$ as $\epsilon_1 \rightarrow 0$. After determining l_j , the relaying receiver transmits $\mathbf{x}_{1,j+1}(l_j)$ at the next block $j + 1$.

Decoding at the Relaying Receiver: At the end of each block j , the relaying receiver finds the unique message $\hat{m}_{1j} \in [1 : 2^{nR_1}]$ such that

$$(\mathbf{u}_j(\hat{m}_{1j}), \mathbf{y}_{1j}) \in \mathcal{T}_{\epsilon_2}^{(n)}(U, Y_1), \quad (33)$$

where $\epsilon_2 > \epsilon_1$. If there are no or more than one such messages, declare error.

Decoding at the Second Receiver: At the end of the whole transmission of J blocks, the second receiver finds the unique message $\hat{m}_2 \in [1 : 2^{nJR_2}]$ such that

$$(\mathbf{v}_j(\hat{m}_2), \mathbf{x}_{1j}(\hat{l}_{j-1}), \hat{\mathbf{y}}_{1j}(\hat{l}_j|\hat{l}_{j-1}), \mathbf{y}_{2j}) \in \mathcal{T}_{\epsilon_3}^{(n)}(V, X_1, \hat{Y}_1, Y_2) \quad (34)$$

for all $j \in [1 : J]$ for some $\hat{l}_1, \hat{l}_2, \dots, \hat{l}_J$, where $\epsilon_3 > \epsilon_1$. If there are no or more than one such messages, declare error.

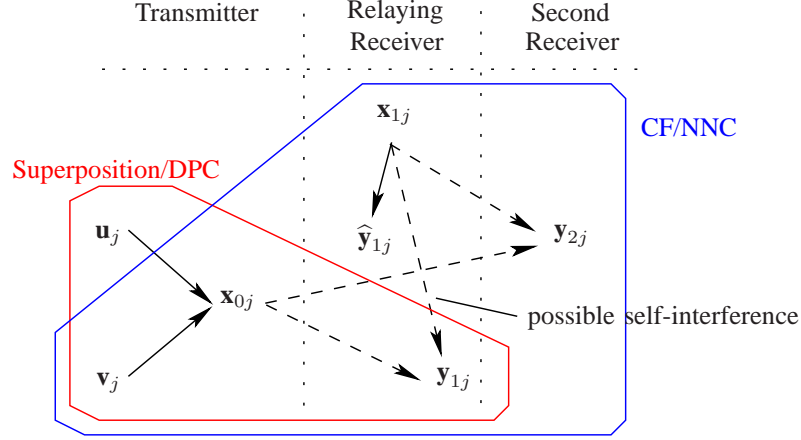


Fig. 5. Markov chain relationship between codewords (solid arrows: codeword Markov chain and dashed arrows: channel links)

Analysis of the Error Probability: Without loss of generality, we assume that truly transmitted message indices are $M_{11} = \dots = M_{1J} = M_2 = 1$ and $L_1 = \dots = L_J = 1$. Then, decoding error occurs only if one or more of the following events occur:

- $\mathcal{E}_1 := \{(\hat{\mathbf{Y}}_{1j}(l_j|1), \mathbf{X}_{1j}(1), \mathbf{Y}_{1j}) \notin \mathcal{T}_{\epsilon_1}^{(n)} \text{ for all } l_j \text{ for some } j \in [1 : J]\}$.
- $\mathcal{E}_2 := \{(\mathbf{U}_j(1), \mathbf{Y}_{1j}) \notin \mathcal{T}_{\epsilon_2}^{(n)} \text{ for some } j \in [1 : J]\}$.
- $\mathcal{E}_3 := \{(\mathbf{U}_j(m_{1j}), \mathbf{Y}_{1j}) \in \mathcal{T}_{\epsilon_2}^{(n)} \text{ for some } m_{1j} \neq 1 \text{ and for some } j \in [1 : J]\}$.
- $\mathcal{E}_4 := \{(\mathbf{V}_j(1), \mathbf{X}_{1j}(1), \hat{\mathbf{Y}}_{1j}(1|1), \mathbf{Y}_{2j}) \notin \mathcal{T}_{\epsilon_3}^{(n)} \text{ for some } j \in [1 : J]\}$.
- $\mathcal{E}_5 := \{(\mathbf{V}_j(m_2), \mathbf{X}_{1j}(l_{j-1}), \hat{\mathbf{Y}}_{1j}(l_j|l_{j-1}), \mathbf{Y}_{2j}) \in \mathcal{T}_{\epsilon_3}^{(n)} \text{ for all } j \text{ for some } (l_1, \dots, l_J), m_2 \neq 1\}$,

where the notations for typical sets are simplified. By the union bound, the error probability is bound as follows:

$$P(\mathcal{E}) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2 \cap \mathcal{E}_1^c) + P(\mathcal{E}_3 \cap \mathcal{E}_1^c) + P(\mathcal{E}_4 \cap \mathcal{E}_1^c) + P(\mathcal{E}_5). \quad (35)$$

The first term $P(\mathcal{E}_1)$ tends to zero as $n \rightarrow \infty$ by the covering lemma [5] if $\hat{R}_2 > I(\hat{Y}_1; Y_1|X_1) + \delta_1(\epsilon_1)$. The second term $P(\mathcal{E}_2)$ tends to zero as $n \rightarrow \infty$ because $\mathbf{U}_j(1) \rightarrow \mathbf{Y}_{1j}$. The third term

$P(\mathcal{E}_3)$ tends to zero as $n \rightarrow \infty$ by the packing lemma [5] if

$$R_1 < I(U; Y_1). \quad (36)$$

The fourth term $P(\mathcal{E}_4 \cap \mathcal{E}_1^c)$ tends to zero as $n \rightarrow \infty$ by the Markov lemma [5], since $(\mathbf{V}_j(1), \mathbf{X}_{1j}(1), \hat{\mathbf{Y}}_{1j}(1|1)) \in \mathcal{T}_{\epsilon_1}^{(n)}$ and

$$\hat{\mathbf{Y}}_{1j} \rightarrow (\mathbf{V}_j, \mathbf{X}_{1j}) \rightarrow \mathbf{Y}_{2j}. \quad (37)$$

Finally, for the fifth term (The proof written in here is similar to that of [12]), define the events

$$\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j) = \{(\mathbf{V}_j(m), \mathbf{X}_{1j}(l_{j-1}), \hat{\mathbf{Y}}_{1j}(l_j|l_{j-1}), \mathbf{Y}_{2j}) \in \mathcal{T}_{\epsilon}^{(n)}\}. \quad (38)$$

Then, we can see that

$$P(\mathcal{E}_5) = P(\cup_{m \neq 1} \cup_{l_1, \dots, l_J} \cap_{j=1}^J \tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \quad (39)$$

$$\leq \sum_{m \neq 1} \sum_{l_1, \dots, l_J} P(\cap_{j=1}^J \tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \quad (40)$$

$$= \sum_{m \neq 1} \sum_{l_1, \dots, l_J} \prod_{j=1}^J P(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \quad (41)$$

$$\leq \sum_{m \neq 1} \sum_{l_1, \dots, l_J} \prod_{j=2}^J P(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)). \quad (42)$$

Now, consider the probability of the event (38). First, assume that $l_{j-1} = 1$. Then, by the joint typicality lemma [5] we have for $l_{j-1} = 1$,

$$P(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) = P\{(\mathbf{V}_j(m_2), \mathbf{X}_{1j}(l_{j-1}), \hat{\mathbf{Y}}_{1j}(l_j|l_{j-1}), \mathbf{Y}_{2j}) \in \mathcal{T}_{\epsilon_3}^{(n)}\} \quad (43)$$

$$\leq 2^{-n(I_1 - \delta_3(\epsilon_3))}, \quad (44)$$

where $I_1 = I(V; \hat{Y}_1, Y_2 | X_1)$, since $\mathbf{V}_j(m_2)$ is independent of $\hat{\mathbf{Y}}_{1j}(l_j|l_{j-1})$ and \mathbf{Y}_{2j} for given $\mathbf{X}_{1j}(l_{j-1})$ due to $M_2 = 1 \neq m_2$. Second, assume that $l_{j-1} \neq 1$. Then, $(\mathbf{V}_j(m_2), \mathbf{X}_{1j}(l_{j-1}), \hat{\mathbf{Y}}_{1j}(l_j|l_{j-1}))$ is independent of \mathbf{Y}_{2j} . Then, by [12, Lemma 2], which is an application of the joint typicality lemma, we have for $l_{j-1} \neq 1$,

$$P(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \leq 2^{-n(I_2 - \delta_3(\epsilon_3))}, \quad (45)$$

where $I_2 = I(V, X_1; Y_2) + I(\hat{Y}_1; V, Y_2|X_1)$. If l_1, l_2, \dots, l_{J-1} have k 1's, then by (44) and (45) we have

$$\prod_{j=2}^n P(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \leq 2^{-n(kI_1 + (J-1-k)I_2 - (J-1)\delta_3(\epsilon_3))}. \quad (46)$$

Therefore, from (42) we have

$$P(\mathcal{E}_5) \leq \sum_{m \neq 1} \sum_{l_1, \dots, l_J} \prod_{j=2}^J P(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \quad (47)$$

$$\leq \sum_{m \neq 1} \sum_{l_J} \sum_{l_1, \dots, l_{J-1}} \prod_{j=2}^b P(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \quad (48)$$

$$\leq \sum_{m \neq 1} \sum_{l_J} \sum_{k=0}^{J-1} \binom{J-1}{k} 2^{n(J-1-k)\hat{R}_2} \cdot 2^{-n(kI_1 + (J-1-k)I_2 - (J-1)\delta_3(\epsilon_3))} \quad (49)$$

$$= \sum_{m \neq 1} \sum_{l_J} \sum_{k=0}^{J-1} \binom{J-1}{k} 2^{-n(kI_1 + (J-1-k)(I_2 - \hat{R}_2) - (J-1)\delta_3(\epsilon_3))} \quad (50)$$

$$\leq 2^{nJR_2} \cdot 2^{n\hat{R}_2} \cdot 2^J \cdot 2^{-n((J-1)\min\{I_1, I_2 - \hat{R}_2\} - (J-1)\delta_3(\epsilon_3))}, \quad (51)$$

which tends to zero as $n \rightarrow \infty$, if

$$R_2 < \frac{J-1}{J}(\min\{I_1, I_2 - \hat{R}_2\} - \delta_3(\epsilon_3)) - \frac{\hat{R}_2}{J}. \quad (52)$$

(In (49), the term $2^{n(J-1-k)\hat{R}_2}$ accounts for the number of $l_{j-1} \neq 1$. Eliminating \hat{R}_2 by substituting $I(\hat{Y}_1; Y_1|X_1) + \delta_1(\epsilon_1)$ from the condition $\hat{R}_2 > I(\hat{Y}_1; Y_1|X_1) + \delta_1(\epsilon_1)$ and sending $J \rightarrow \infty$, we obtain

$$R_2 < \min\{I(V; \hat{Y}_1, Y_2|X_1), I(V, X_1; Y_2) - I(\hat{Y}_1; Y_1|V, X_1, Y_2)\} - \delta_1(\epsilon_1) - \delta_3(\epsilon_3). \quad (53)$$

Since δ_1 and δ_3 converge to zero, we have the claim by (36) and (53). \blacksquare

APPENDIX B

ACHIEVABLE RATE REGION FOR THE RBC-CF/NNC SCHEME IN THE GAUSSIAN CASE

In the Gaussian case, we have $p(u) \sim \mathcal{CN}(0, \alpha P_0)$, $p(v) \sim \mathcal{CN}(0, \bar{\alpha} P_0)$, and $p(x_1) \sim \mathcal{CN}(0, P_1)$. Furthermore, we have (14) and (21) for $p(x_0|u, v)$ and $p(\hat{y}_1|y_1, x_1)$, respectively.

We need to compute R_1 and R_2 in (9) and (10) based on (14), (21), (12), and (13). Since

$$Y_1 = h_{01}(U + V) + Z_1 \quad (54)$$

$$\hat{Y}_1 = h_{01}V + Z_1 + \hat{Z} \quad (55)$$

$$Y_2 = h_{02}(U + V) + h_{12}X_1 + Z_2, \quad (56)$$

the achievable rate region in Theorem 3 is given by

$$\begin{aligned} R_1 &< I(U; Y_1) \\ &= I(U; h_{01}U + h_{01}V + Z_1) \end{aligned} \quad (57)$$

$$\begin{aligned} R_2 &< \min\{I(V; \hat{Y}_1, Y_2 | X_1), I(V, X_1; Y_2) - I(\hat{Y}_1; Y_1 | V, X_1, Y_2)\} \\ &= \min\{I(V; h_{01}V + Z_1 + \hat{Z}, h_{02}V + h_{02}U + Z_2), \\ &\quad I(V, X_1; h_{02}V + h_{12}X_1 + h_{02}U + Z_2) - I(Z_1 + \hat{Z}; h_{01}U + Z_1 | h_{02}U + Z_2)\} \end{aligned} \quad (58)$$

Then, the term in (57) and the first argument of the minimum in (58) are respectively given by

$$I(U; h_{01}U + h_{01}V + Z_1) = \log \left(1 + \frac{|h_{01}|^2 \alpha P_0}{|h_{01}|^2 \bar{\alpha} P_0 + N_1} \right) \quad (59)$$

$$I(V; h_{01}V + Z_1 + \hat{Z}, h_{02}V + h_{02}U + Z_2) = \log \left(1 + \frac{|h_{01}|^2 \bar{\alpha} P_0}{N_1 + \hat{N}} + \frac{|h_{02}|^2 \bar{\alpha} P_0}{|h_{02}|^2 \alpha P_0 + N_2} \right) \quad (60)$$

The first term of the second argument in the minimum in (58) is expressed as

$$I(V, X_1; h_{02}V + h_{12}X_1 + h_{02}U + Z_2) = \log \left(1 + \frac{|h_{02}|^2 \bar{\alpha} P_0 + |h_{12}|^2 P_1}{|h_{02}|^2 \alpha P_0 + N_2} \right). \quad (61)$$

Finally, the second term of the second argument in the minimum in (58) can be expressed as

$$\begin{aligned}
& I(Z_1 + \hat{Z}; h_{01}U + Z_1 | h_{02}U + Z_2) \\
&= h(Z_1 + \hat{Z} | h_{02}U + Z_2) - h(Z_1 + \hat{Z} | h_{01}U + Z_1, h_{02}U + Z_2) \\
&= h(Z_1 + \hat{Z}) - h(Z_1 + \hat{Z} | h_{01}U + Z_1, h_{02}U + Z_2) \\
&= \log(N_1 + \hat{N}) - \log\left(\frac{N_2|h_{01}|^2\alpha P_0 + N_1|h_{02}|^2\alpha P_0 + N_1N_2}{N_1N_2|h_{01}|^2\alpha P_0 + \hat{N}N_2|h_{01}|^2\alpha P_0 + \hat{N}N_1|h_{02}|^2\alpha P_0 + \hat{N}N_1N_2}\right) \\
&= \log\left(1 + \frac{N_1^2N_2 + N_1^2|h_{02}|^2\alpha P_0}{\hat{N}N_1N_2 + \hat{N}N_2|h_{01}|^2\alpha P_0 + \hat{N}N_1|h_{02}|^2\alpha P_0 + N_1N_2|h_{01}|^2\alpha P_0}\right). \tag{62}
\end{aligned}$$

REFERENCES

- [1] Q. Li, H. Niu, A. Papathanassiou, and G. Wu, “5G Network Capacity: Key Elements and Technologies,” *IEEE Veh. Technol. Mag.*, vol. 9, pp. 71 – 78, Mar. 2014.
- [2] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access,” in *Proc. IEEE VTC Spring*, 2013.
- [3] Z. Ding, M. Peng, and H. V. Poor, “Cooperative Non-Orthogonal Multiple Access in 5G Systems,” *IEEE Commun. Letter*, vol. 19, pp. 1462 – 1465, Aug. 2015.
- [4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [5] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. New York: Cambridge University Press, 2011.
- [6] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, “Device-to-Device Communication as an Underlay to LTE-Advanced Networks,” *IEEE Commun. Mag.*, vol. 7, pp. 42 – 49, Dec. 2009.
- [7] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, “Design Aspects of Network Assisted Device-to-Device Communications,” *IEEE Commun. Mag.*, vol. 50, pp. 170 – 177, Mar. 2012.
- [8] Y. Liang and V. V. Veeravalli, “The Impact of Relaying on the Capacity of Broadcast Channels,” in *Information Theory Proceedings (ISIT), 2004 IEEE International Symposium on*, (Chicago, USA), pp. 403 – 403, Jun. 2004.
- [9] S. I. Bross, “On the Discrete Memoryless Partially Cooperative Relay Broadcast Channel and the Broadcast Channel with Cooperative Decoders,” *IEEE Trans. Inform. Theory*, vol. 55, pp. 2161 – 2182, May. 2004.
- [10] T. M. Cover and A. E. Gamal, “Capacity Theorems for the Relay Channel,” *IEEE Trans. Inform. Theory*, vol. 25, pp. 572 – 584, Sep. 1979.
- [11] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative Strategies and Capacity Theorems for Relay Networks,” *IEEE Trans. Inform. Theory*, vol. 51, pp. 3037 – 3063, Sep. 2005.

- [12] S. H. Lim, Y.-H. Kim, A. E. Gamal, and S.-Y. Chung, “Noisy Network Coding,” *IEEE Trans. Inform. Theory*, vol. 57, pp. 3132 – 3152, May. 2011.
- [13] M. H. M. Costa, “Writing on Dirty paper,” in *IEEE Trans. Inform. Theory*, vol. 29, pp. 439 – 441, May. 1979.
- [14] S. I. Gel’fand and M. S. Pinsker, “Coding for Channel with Random Paramters,” *Probl. Contr. Inform. Theory*, vol. 9, pp. 19 – 31, Jan. 1980.
- [15] A. Benjebbour, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, “System-Level Performance of Downlink NOMA for Future LTE Enhancements,” in *Proc. IEEE Globecom 2013*, (Atlanta, USA), Dec. 2013.
- [16] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic Beamforming Using Dumb Antennas,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277 – 1294, Jun. 2002.
- [17] H. Holma and A. Toskala, *WCDMA for UMTS*. New York: Wiley, 2001.
- [18] U. Erez and S. ten Brink, “A Close-to-Capacity Dirty Paper Coding Scheme,” *IEEE Trans. Inform. Theory*, vol. 51, pp. 3417 – 3432, Oct. 2005.
- [19] Y. Sun, Y. Yang, A. D. Liveris, V. Stanković, and Z. Xiong, “Near-Capacity Dirty-Paper Code Design: A Source-Channel Coding Approach,” *IEEE Trans. Inform. Theory*, vol. 55, pp. 3013 – 3031, Jul. 2009.
- [20] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, “Superposition Coding for Side-Information Channels,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 1872 – 1889, May. 2006.
- [21] R. B.-Serrano, R. Thobaben, V. Rathi, and M. Skoglund, “Polar Codes for Compress-and-Forward in Binary Relay Channels,” in *Proc. Asilomar Conf. Signals, Systems and Computers, Pacific Grove, CA. USA*. Nov. 2010.
- [22] R. B.-Serrano, “Coding Strategies for Compress-and-Forward Relaying,” *Licentiate Thesis, Royal Institutue of Technology (KTH), Stockholm, Sweden*, 2010.
- [23] R. B.-Serrano, R. Thobaben, M. Andersson, V. Rathi, and M. Skoglund, “Polar Codes for Cooperative Relaying,” *IEEE Trans. Commun.*, vol. 60, pp. 3263 – 3273, Nov. 2012.
- [24] M. Karzand, “Polar Codes for Degraded Relay Channels,” in *Proc. Int. Zurich Seminar Commun., Zurich, Switzerland*, Feb. 2012.
- [25] D. S. Karas, K. N. Pappi, and G. K. Karagiannidis, “Smart Decode-and-Forward Relaying with Polar Codes,” *IEEE Commun. Letter*, vol. 3, pp. 62 – 65, Feb. 2014.